



Revisiting the Prediction of the Partition Coefficient Using Variables Determined by the Conductor-Like Polarizable Continuum Model

José G. Parra 

Laboratory of Molecular Simulation for Experimental Chemistry (SIMOLQUIMEX), Department of Chemistry, Faculty of Experimental Sciences and Technology, University of Carabobo, Chemistry Building, Salvador Allende Avenue, Bárbula, Venezuela

Article History

Received: May 01, 2026

Accepted: July 22, 2026

Published: August 25, 2026

Abstract

The Conductor-like Polarizable Continuum Model (C-PCM), together with the Amovilli-Mennucci method implemented in the GAMESS-US package, was used to estimate the electrostatic, dispersion, and repulsion free energies of organic molecules in 1-octanol and water. Solvation calculations in water and 1-octanol were performed using the C-PCM/PBE-VWN/6-31G(2p,2d) method to determine the electrostatic contribution to the solvation free energy. A quantitative structure–property relationship (QSPR) approach was then used to derive a model for estimating the cavitation free energies of organic molecules from their experimental solvation free energies. In addition, using the atomic surface areas of the C, H, N, and O atoms in the organic molecules, calculated with the GEPOL-GB method, along with the electrostatic and non-bonding free energies, two QSPR models were constructed to predict the 1-octanol/water partition coefficient ($\log P$) for neutral organic molecules containing a single functional group from the following classes: alcohols, ketones, aldehydes, carboxylic acids, esters, ethers, amines, alkanes (linear and cyclic), and aromatic compounds. For alcohols, 1,2-ethanediol, 1,3-propanediol, and 1,4-butanediol were included in the training set. Model A, which uses electrostatic free energies and atomic surface areas, achieved a coefficient of determination of $R^2 = 0.991$, while Model B, which incorporates non-bonding free energies, achieved $R^2 = 0.986$. The applicability domain of the models is explicitly defined for monofunctional compounds at infinite dilution and 298.15 K.

Keywords:

electrostatic free energy; partition coefficient; solvation; computational chemistry; density functional theory

1. Introduction

The equilibrium between liquid phases, or between liquid and vapor, is critical to understanding a wide range of chemical and biological phenomena; indeed, understanding and predicting such equilibria is essential for optimizing chemical products and processes [1]. Researchers have generally employed these types of equilibria to theoretically estimate a range of properties, including the aqueous solvation energy of polar and nonpolar species, the vapor pressure and boiling point of pure liquids, the equilibrium constant of carboxylic acids in the aqueous

phase, and the partition coefficient of a solute between two immiscible liquids [2–8].

The partition coefficient of a solute between two immiscible liquids is one of the most significant properties used in this context. This property is particularly valuable for describing the behavior of amphiphilic molecules and species used in the pharmaceutical industry. The distribution of a molecule between water and a lipid material is typically described using the 1-octanol/water partition coefficient $K_{o/w}$, commonly expressed as ($K_{o/w} = \log P$), which is regarded as the most important physicochemical

Corresponding Author:

José G. Parra, Laboratory of Molecular Simulation for Experimental Chemistry (SIMOLQUIMEX), Department of Chemistry, Faculty of Experimental Sciences and Technology, University of Carabobo, Chemistry Building, Salvador Allende Avenue, Bárbula, Venezuela, jgparra2@uc.edu.ve



© 2026 Copyright by the Author.

Licensed as an open access article using a **CC BY 4.0 license**.

property for describing the affinity of chemical species for a given solvent [9–11].

1-Octanol is widely accepted as the material whose behavior most closely resembles that of lipids found in nature, and the partition coefficient is the most widely used descriptor in structure-property relationship (QSPR) studies [12]. Concurrently, QSPR models emerge from a process that begins with the description of molecular structures and culminates in the prediction of molecular behavior across different phases [13–16]. Consequently, QSPR models have been developed using properties calculated by quantum mechanical methods and molecular dynamics [17–20].

A variety of quantum mechanical solvation models, including C-PCM (Conductor-Like Polarizable Continuum Model) and COSMO (Conductor-Like Screening Model), have been used to investigate and calculate the properties of chemical compounds in solvents [21–27]. In these methods, the solvent is represented by a continuous dielectric model, which provides a satisfactory description of molecular properties at low computational cost and yields favorable results. Numerous methods have also been developed to predict the molecular properties of solutes in different solvent media, including the molecular volume of liquids, vapor pressure, and solvation energies [28–32]. Accordingly, this study employs the C-PCM solvation method to calculate the structural and energetic properties of a set of organic molecules. Two quantitative structure-property relationship models were then obtained to estimate the 1-octanol/water partition coefficient ($K_{o/w} = \log P$) and the free energies of solvation for these groups of organic molecules. While solvation free energies are essential for understanding solute-solvent interactions, the partition coefficient is inherently a transfer property between two phases. The distinction between transfer and solvation free energies is critical because the former accounts for the competitive solvation of the solute by two immiscible solvents, including effects of activity coefficients, conformational changes, and specific solvent-solvent interactions that are not captured by single-solvent continuum models.

2. Theoretical Framework

The solvation free energy (ΔG_{solv}) is the free energy change required to transfer a molecule from the vacuum into the solvent [23] and can be defined by Equation (1):

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{disp} + \Delta G_{rep} + \Delta G_{cav} + RT \ln \left(\frac{q_{rot,g} q_{vib,g}}{q_{rot,s} q_{vib,s}} \right) - RT \ln \left(\frac{n_{M,g} \Lambda_{M,g}^3}{n_{M,s} \Lambda_{M,s}^3} \right) \quad (1)$$

The term ΔG_{elec} is the electrostatic component arising from the interaction between the solute and the solvent medium. The variables ΔG_{rep} and ΔG_{disp} are the free energies of repulsion and dispersion between the solute and solvent. Finally, the term ΔG_{cav} is the free energy required to construct the cavity of the solute in the solvent; the last two terms are negligible.

2.1. Electrostatic Free Energy

Under the influence of a solvent described by a dielectric continuum model, the Hamiltonian of the solute can be defined by Equation (2) as:

$$\hat{H} = \hat{H}^o + \hat{V} \quad (2)$$

Here, $\hat{H}^o$ is the Hamiltonian of the isolated solute, and the operator $\hat{V}$ describes the solute-solvent interaction, which depends linearly on the state function Ψ of the solute. In this case, the electrostatic free energy ΔG_{elec} is determined by variational minimization of the free energy of the solute, given by Equation (3):

$$\Delta G_{elec} = \langle \Psi | \hat{H}^o | \Psi \rangle + \frac{1}{2} \langle \Psi | \hat{V} | \Psi \rangle \quad (3)$$

In general, the interaction operator $\hat{V}$ is written in terms of apparent polarization charges. In each tessera i , a charge q_i appears, according to the conductor-like boundary condition:

$$V\left(\vec{r}\right) + \sum_i^{segments} V_{q_i}\left(\vec{r}\right) = 0 \quad (4)$$

where $V\left(\vec{r}\right)$ and $V_{q_i}\left(\vec{r}\right)$ are the electrostatic potentials due to the solute and to the polarization charges, respectively, and $\vec{r}$ is a point on the segment surface.

The vector of polarization charges Q can be defined by Equation (5):

$$AQ = -V \quad (5)$$

Here, the vector V contains the electrostatic potential due to the solute on the surface segment. The elements of matrix A are expressed by Equations (6) and (7):

$$A_{ii} = 1.07 \sqrt{\frac{4\pi}{S_i}} \quad (6)$$

$$A_{ij} = \frac{1}{\left| \vec{r}_i - \vec{r}_j \right|} \quad (7)$$

where S_i is the area of tessera i . Equation (6) was defined by Klamt and Schüürmann for a simple sphere partitioned

into segments of identical surface area [21]. In C-PCM, the polarization charges are scaled so that the total polarization charge obeys Gauss's law; an effective way to achieve this is to multiply each charge by the factor $\frac{\varepsilon-1}{\varepsilon}$, yielding new charges given by Equation (8):

$$q = \frac{\varepsilon - 1}{\varepsilon} Q \quad (8)$$

Here, ε corresponds to the dielectric constant of the solvent. At the same time, the contribution ΔG_{elec} is important for polar and charged solutes because of the polarization of the solvent, which is modeled as a uniform medium of a given dielectric constant.

2.2. Free Energy Required for the Construction of a Molecular Cavity in the Solvent

To determine solute-solvent interactions, the molecules are placed in cavities formed by overlapping spheres centered on the solute atoms. Constructing the cavity for the solute in the solvent requires a certain amount of energy, known as the cavitation free energy, a concept introduced by Uhlig [33] and initially applied to determine the solubility of gases in liquids. To determine the cavitation free energy, ΔG_{cav} , several models have been developed, including the formulation proposed by Reiss and Pierotti [34–36]. This theory describes fluid molecules using hard spheres to evaluate ΔG_{cav} , which is determined by a polynomial depending on the radius R_{ms} , corresponding to the sum of the radii of the solute and solvent:

$$\Delta G_{cav} = K_0 + K_1 R_{ms} + K_2 R_{ms}^2 + K_3 R_{ms}^3 \quad (9)$$

The coefficients K_i ($i = 0, 1, 2, 3$) are expressed in terms of the properties of the solvent, including its molecular radius and number density, and those of the system (pressure P and temperature T). These can be calculated using Equations (10)–(13):

$$K_0 = RT \left[-\ln(1-y) + \frac{9}{2} \left(\frac{y}{1-y} \right)^2 \right] - \frac{4\pi R_s^3 P}{3} \quad (10)$$

$$K_1 = -\frac{3RT}{R_s} \left[\frac{y}{1-y} + 3 \left(\frac{y}{1-y} \right)^2 \right] + 4\pi R_s^2 P \quad (11)$$

$$K_2 = \frac{3RT}{R_s^2} \left[\frac{y}{1-y} + \frac{3}{2} \left(\frac{y}{1-y} \right)^2 \right] - 4\pi R_s P \quad (12)$$

$$K_3 = \frac{4\pi P}{3} \quad (13)$$

The term $y = \frac{4\pi R_s^3 n_s}{3}$ is associated with the microscopic terms. The cavitation free energy is a positive term that accounts for the loss of entropy of the solvent due to the reorganization of its molecules in the presence of the solute.

2.3. Free Energies of Dispersion and Repulsion

This section presents the models used to calculate the dispersive and repulsive contributions to the solvation free energy. Using the atom-atom interaction model [37], the dispersion free energy between two molecules A and S is described by Equation (14):

$$E(A, S)_{disp} = -\sum_{n=6}^{\infty} \sum_{t \in S} \sum_{a \in A} d_{at}^{(n)} r_{at}^{-n} \quad (14)$$

Here, $d_{at}^{(n)}$ is the dispersion coefficient and r_{at} is the distance between atoms. In real solutions, the description of the relative distances between the atoms of the solute and solvent can be replaced by the distribution function defined in Equation (15):

$$\rho_{at}(r_{at}) = N_t \rho_s g_{at}(r_{at}) \quad (15)$$

In Equation (15), N_t is the number of atoms of type t for each solvent molecule, ρ_s is the number density of the solvent, and g_{at} is the correlation function, which depends on the interaction distance (r_{at}). Using a continuous function allows the summation over S to be replaced by an integral, shown in Equation (16):

$$G_{disp}(AenS) = -\sum_{t \in S} \sum_{a \in A} d_{at}^{(6)} N_t \rho_s \times \int dr_t g_{at}(r_{at}) r_{at}^{-6} \quad (16)$$

Using Equation (16), it is possible to determine the dispersion free energy of a solute in a solvent. In turn, Amovilli [38] proposed a model to calculate the dispersion free energy using a quantum-mechanical treatment:

$$\Delta G_{disp} = \frac{1}{\pi} \int_0^\infty d\omega \sum_{K \neq 0} \frac{\omega_K^A}{\omega_{0K}^A \omega_K^2 + \omega^2} \times \int d\vec{r}_1 \oint \frac{d\vec{r}_2}{r_{12}^2} P_A(0K|\vec{r}_1) \sigma_B[B(i\omega), P_A(0K|\vec{r})] (\vec{r}_2) \quad (17)$$

where $P_A(0K|\vec{r})$ and ω_{0K}^A are the densities and energies of transition of a state K of the solute. The term σ_B is the surface charge density induced by the electric field of the charge distribution $P_A(0K|\vec{r})$, which depends on a dielectric constant calculated at imaginary frequencies. Additionally, σ_B is always smaller than the square of the refractive index η_B^2 , estimated at zero frequency and measured from the electronic transitions in the visible spectrum. However, Mennucci and Amovilli [39] defined these terms using Equation (18):

$$\sigma_B[B(i\omega), P_A(0K)]\left(\vec{r}\right) = \frac{\Omega_B^2}{\Omega_B^2 + \omega^2} \frac{\eta_B^2 - 1}{4\pi\eta_B^2} E_{0K}^A\left(\vec{r}\right) \quad (18)$$

where $\Omega_B = \eta_B I_B$ depends on the ionization potential I_B of the solvent. Here, $E_{0K}^A(\vec{r})$ is the component of the electric field orthogonal to the surface of the solute cavity at the point $\vec{r}$, associated with the transition density $P_A(0K|\vec{r})$. Using Equations (18) and (17), the dispersion free energy can be defined as:

$$\Delta G_{disp} = - \frac{\eta_B^2 - 1}{8\pi\eta_B \left(\eta_B + \frac{\omega_A}{I_B}\right)} \sum_{K \neq 0} \oint d\vec{r} V_{0K}^A(\vec{r}) E_{0K}^A(\vec{r}) \quad (19)$$

The term V_{0K}^A is the potential associated with the electric field E_{0K}^A . All terms associated with the dispersion free energy equation are calculated using the restricted Hartree-Fock method.

To calculate the repulsion free energy in the solvation process, Amovilli and Mennucci [39] proposed a model based on the method of Floris [37] and Equation (20):

$$\Delta G_{rep} = \rho_B \int d\vec{R} U_{rep}^{AB}(\vec{R}) g_{AB}(\vec{R}) \quad (20)$$

Here, A and B correspond to the solute and the solvent, respectively. $\vec{R}$ is an appropriate set of coordinates defining the internal geometry of the complex AB , ρ_B is the number density of the solvent, and g_{AB} is a correlation function whose value is zero inside the solute cavity and one outside it. This model is based on the principle that the repulsive force between two interacting molecules follows the Pauli exclusion principle. Based on this principle, the free energy of repulsion increases with the overlap of the two distributions and is related to the density of electrons with the same spin. Thus, the repulsion free energy in the solvation process can be calculated using Equation (21):

$$\Delta G_{rep} = \rho_B n_{par}^B \int \vec{R} \frac{1}{2} \int \frac{d\vec{r}_1 d\vec{r}_2}{r_{12}} P_A(\vec{r}_1, \vec{r}_2) P_{par}\left(\vec{R} \left| \vec{r}_2, \vec{r}_1 \right. \right) \quad (21)$$

Here, n_{par}^B is the number of pairs of valence electrons of the solvent, and $P_{par}(\vec{R}|\vec{r}_2, \vec{r}_1)$ and $P_A(\vec{r}_1, \vec{r}_2)$ correspond to the adapted pair density matrices. Here, $\vec{R}$ are the coordinates of the center of the localized orbital containing the pairs of free or bonded electrons of the solute molecule.

2.4. Distribution of a Solute in Two Immiscible Liquids

Consider two immiscible liquids in contact, into which a non-dissociable solute A is incorporated at a concentration within the dilute-solution range [9,40]. The solute A distributes between the two phases to form a solution. If both phases are in equilibrium, the chemical potentials of solute A in both phases are equal and can be written as:

$$\mu_{(solute,\alpha)} = \mu_{(solute,\beta)} \quad (22)$$

In turn, the chemical potential of solute A in a given phase can be written as:

$$\mu_{(solute,i)} = \mu_{(solute,i)}^0 + RT \ln x_{(solute,i)} \quad (23)$$

where $x_{(solute,i)}$ corresponds to the fraction of solute in equilibrium with one of the phases, and $\mu_{(solute,i)}^0$ is the standard potential of solute A in one of the phases. Using Equations (22) and (23), a relationship can be established to calculate the solute distribution between the two phases:

$$\mu_{(solute,\alpha)}^0 + RT \ln x_{(solute,\alpha)} = \mu_{(solute,\beta)}^0 + RT \ln x_{(solute,\beta)} \quad (24)$$

where the concentration ratio is given by:

$$\left(\frac{x_{(solute,\alpha)}}{x_{(solute,\beta)}} \right) = K_{distribution} = EXP \left(\frac{\mu_{(solute,\beta)}^0 - \mu_{(solute,\alpha)}^0}{RT} \right) \quad (25)$$

Assuming that $\mu_{(solute,i)}^0 = \Delta G_{(solv,i)}^0 + \Delta G_{vap(A)}^0$, where $\Delta G_{vap(A)}^0$ is the vaporization free energy of the solute, we can calculate the solvation of A in a given solvent. In this case, the standard chemical potential of solute A in a phase $\mu_{(solute,i)}^0$ is equivalent to the dissolution free energy ΔG_{sol}^0 . Therefore, the distribution constant of the solute in both phases can be written as follows:

$$EXP \left(\frac{\Delta G_{vap}^0(A) + \Delta G_{(solv,\beta)}^0 - (\Delta G_{vap}^0(A) + \Delta G_{(solv,\alpha)}^0)}{RT} \right) \quad (26)$$

Here, $\Delta G_{(solv,\beta)}^0$ and $\Delta G_{(solv,\alpha)}^0$ correspond to the solvation free energies of the solute in the two immiscible liquids. Simplifying Equation (26) yields an expression, based on quantum mechanical solvation methods, for calculating the distribution constant $K_{\frac{o}{w}}$ of a solute A between two immiscible liquids:

$$K_{o/w} = EXP \left(\frac{\Delta G_{(solv,\beta)}^0 - \Delta G_{(solv,\alpha)}^0}{RT} \right) \quad (27)$$

Generally, it is possible to calculate the free energies of phase changes and the solvation free energy. However, determining the free energy associated with the transfer of a solute from one liquid phase to another remains a limitation of quantum mechanical methods used to estimate the chemical potential of a system. Therefore, Equation (26) is particularly suitable for estimating the free energy associated with this phase transfer.

3. Methodology

3.1. Quantum Mechanical Calculations

All three-dimensional molecular structures were constructed using the wxMacMolPlt v7.4.5 software [41]. The molecular structures were then optimized in vacuum using density functional theory (DFT), with the PBE-VWN functional and the 6-31G(2p,2d) basis set [42–44].

Using the coordinates obtained from the optimization, solvation calculations in water and 1-octanol were performed using the C-PCM//PBE-VWN/6-31G(2p,2d) method to determine the electrostatic contribution to the solvation free energy. These calculations were based on the theory of Ben-Naim [45], which assumes that a molecule does not change its structural conformation when transferred from vacuum to the solvent medium. Figure 1 shows the solvation process of ethanol in a continuum dielectric medium according to the Ben-Naim principle.

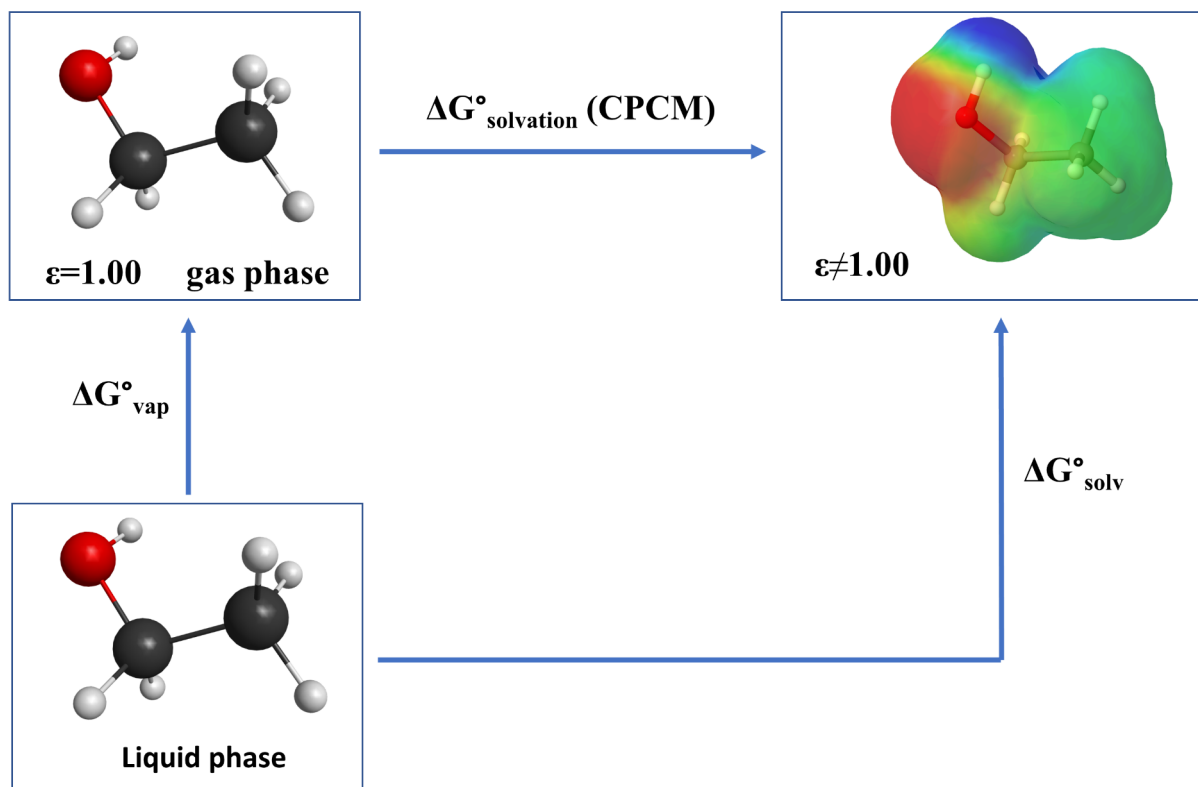


Figure 1: Solvation process of ethanol in a continuum dielectric medium.

To construct the cavities of the solute molecules, the GEPOLE-GB method proposed by Pascual-Ahuir and Tomasi [46,47] was used, together with van der Waals radii for the C, N, H, and O atoms; the radii used are shown in Table 1. For the solvation studies in water and 1-octanol, the dielectric constant, molecular volume, and surface tension were used as input parameters [48,49]. The repulsion and dispersion free energies were calculated using the Amovilli-Mennucci method, which required defining several solvent parameters, including the ionization potential, the square of the refractive index, the density, and the number of valence-shell electrons of the solvent. The ionization potentials of water and 1-octanol were calculated using an SCF//UHF/PBE-VWN optimization procedure with a 6-31G(2p,2d) basis set. All calculations were performed using the GAMESS-US software [50]. The parameters used are shown in Table 2.

Table 1: vdW radii used in the GEPOLE-GB method.

Atom	vdW Radii Å
C	1.70
N	1.60
O	1.50
H	1.20

Table 2: Parameter values used in the Amovilli-Mennucci and C-PCM methods.

Parameter Input	1-Octanol	Water
Dielectric constant (ϵ)	9.86	78.39
Molecular volume (cm^3/mol)	158.23	18.07
Surface tension (mN/m)	27.10	71.81
Ionization potential (Hartree)	0.332	0.396
Square of the refractive index	2.04	1.75

3.2. Training Set Selection and Model Validation

The training set of 42 organic molecules used to parameterize the cavitation free energy model (see Table 3) and the set of 50 molecules used for the QSPR models (Models A and B) were selected based on three primary criteria: (i) availability of high-quality experimental solvation free energies and logP values in the literature; (ii) structural and functional diversity, including alcohols, ketones, aldehydes, carboxylic acids, esters, ethers, amines, alkanes (linear and cyclic), and aromatic compounds; and (iii) variation in molecular size and alkyl chain length, to systematically capture hydrophobic contributions. Selecting a diverse training set is essential for developing QSPR models with broad applicability and reliable predictions for unseen compounds.

Table 3: Electrostatic, repulsion, and dispersion free energies calculated using the C-PCM and Amovilli-Mennucci methods.

	1-Octanol				Water			
	ΔG_{elec} kJ mol^{-1}	ΔG_{disp} kJ mol^{-1}	ΔG_{rep} kJ mol^{-1}	ΔG_{exp} (solv.) kJ mol^{-1}	ΔG_{elec} kJ mol^{-1}	ΔG_{disp} kJ mol^{-1}	ΔG_{rep} kJ mol^{-1}	ΔG_{exp} (Solv.) kJ mol^{-1}
acetone	-16.32	-42.01	9.33	-13.18	-18.62	-37.99	11.63	-16.11
2-butanone	-15.02	-50.46	11.63	-15.82	-17.16	-45.61	14.52	-15.23
2-pentanone	-15.02	-59.46	14.14	-18.20	-17.16	-53.77	17.66	-14.77
3-pentanone	-13.93	-59.08	14.06	-18.24	-15.94	-53.43	17.57	-14.27
2-hexanone	-15.23	-68.29	16.53	-21.00	-17.41	-61.76	20.63	-13.77
1-propanal	-14.06	-42.26	9.41	-17.28	-16.03	-38.20	11.76	-14.39
1-butanal	-14.02	-51.17	11.97	-19.33	-15.94	-46.28	14.94	-13.31
acetic acid	-18.45	-38.49	6.61	-26.57	-20.71	-34.81	8.24	-28.03
propanoic acid	-17.74	-47.28	9.04	-28.70	-19.92	-42.76	11.26	-27.07
butanoic acid	-17.36	-56.11	11.38	-31.72	-19.50	-50.75	14.18	-26.61
pentanoic acid	-17.20	-64.98	13.98	-34.39	-19.29	-58.75	17.45	-25.77
hexanoic acid	-17.32	-73.73	16.19	-36.90	-19.41	-66.70	20.21	-25.98
benzene	-5.10	-51.42	11.59	-15.57	-5.73	-50.96	14.44	-3.64
toluene	-4.98	-59.96	13.93	-19.04	-5.56	-54.23	17.36	-3.72
ethylbenzene	-5.31	-67.91	16.40	-21.26	-5.94	-61.42	20.46	-3.35
methanol	-12.80	-26.57	5.61	-15.82	-14.35	-24.02	6.99	-21.38
ethanol	-12.39	-36.15	8.28	-18.24	-13.93	-32.68	10.33	-20.96
1-propanol	-11.97	-45.23	10.75	-21.00	-13.43	-40.88	13.43	-20.21
1-butanol	-12.26	-54.10	13.22	-23.89	-12.30	-53.60	16.53	-19.75
1-pentanol	-12.64	-63.01	15.57	-26.78	-14.18	-56.95	19.46	-18.70

Table 3: *Cont.*

	1-Octanol				Water			
	ΔG_{elec} kJ mol ⁻¹	ΔG_{disp} kJ mol ⁻¹	ΔG_{rep} kJ mol ⁻¹	ΔG_{exp} (solv.) kJ mol ⁻¹	ΔG_{elec} kJ mol ⁻¹	ΔG_{disp} kJ mol ⁻¹	ΔG_{rep} kJ mol ⁻¹	ΔG_{exp} (Solv.) kJ mol ⁻¹
1-hexanol	-12.68	-72.05	18.12	-29.54	-14.23	-65.15	22.59	-18.24
1-heptanol	-12.85	-80.96	20.54	-32.43	-14.44	-73.22	25.69	-17.74
2-propanol	-11.92	-44.44	10.88	-19.33	-13.43	-40.17	13.56	-19.92
1,2-ethanediol	-22.13	-42.26	7.99	-31.13	-24.85	-38.20	9.96	-38.91
dimethyl ether	-7.28	-35.65	8.66	-8.62	-8.20	-32.22	10.84	-8.03
diethyl ether	-6.28	-54.52	14.14	-12.09	-7.11	-49.33	17.66	-7.36
propyl methyl ether	-6.86	-54.06	13.93	-15.19	-7.74	-48.91	17.41	-6.95
methyl acetate	-14.94	-46.95	9.41	-14.81	-16.82	-42.47	11.72	-13.89
ethyl acetate	-15.06	-56.36	11.92	-16.99	-16.95	-50.96	14.90	-12.97
methyl propionate	-13.98	-55.61	11.80	-16.99	-15.73	-50.29	14.73	-12.26
methyl butanoate	-13.85	-64.86	14.31	-19.21	-15.61	-58.62	17.91	-11.84
cyclopentane	-0.54	-51.30	13.60	-11.09	-0.59	-46.40	16.95	5.02
cyclohexane	-0.46	-58.54	15.52	-14.48	-0.54	-52.93	19.37	5.15
methylcyclohexane	-0.46	-67.11	18.28	-16.23	-0.50	-60.67	22.80	7.15
n-pentane	-0.67	-57.41	16.11	-10.25	-0.75	-51.93	20.13	9.75
n-hexane	-0.75	-66.11	18.28	-12.59	-0.84	-59.79	22.80	10.42
n-heptane	-0.84	-75.02	20.63	-15.65	-0.92	-67.83	25.73	10.96
n-octane	-0.79	-84.14	23.47	-17.49	-0.88	-76.11	29.29	12.09
n-propylamine	-10.46	-46.65	11.97	-19.96	-11.76	-42.18	14.94	-17.53
n-butylamine	-10.50	-55.52	14.48	-22.30	-11.80	-50.21	18.08	-17.95
diethylamine	-6.40	-55.90	14.98	-19.87	-7.20	-50.55	18.70	-17.03
dipropylamine	-6.44	-73.68	19.83	-25.19	-7.24	-66.65	24.77	-15.31

As an exception, 1,2-ethanediol was included in the training set for the cavitation free energy model. In the logP models, by contrast, alcohol molecules such as 1,2-ethanediol, 1,3-propanediol, and 1,4-butanediol were included in the training set to extend the model's applicability, and glycerol was used to validate this extension.

To prevent overfitting and evaluate the predictive power of the models, two complementary validation strategies were employed. First, leave-one-out cross-validation was performed on the 42 molecules used to parameterize the cavitation free energy model, yielding coefficients of determination of $R^2 = 0.985$ for water and $R^2 = 0.982$ for 1-octanol, confirming that the models are not overfitted to the training data. Second, external validation was conducted using an independent set of ten molecules (discussed in Section 4.3) not included in the training set, comprising 2-pentanol, n-hexanal, hexylamine, dibutyl ether, propyl acetate, n-butylbenzene, dioxane, anisole, glycerol, and 3-pentanone. The excellent agreement between predicted and experimental logP values for these molecules (discussed in Section 4) provides strong evidence that the models possess genuine predictive capability and are not the result of overfitting. Additionally, all coefficients in the models (see Sections 4.1 and 4.2) were evaluated for chemical consistency. Positive coefficients were associated with hydrophobic contributions

(carbon surfaces and alkyl groups), whereas negative coefficients were associated with polar contributions (oxygen surfaces and hydrogen-bonding groups), confirming that the models capture meaningful structure-property relationships rather than spurious correlations.

4. Results and Discussion

4.1. Linear Regression Model to Estimate the Cavitation Free Energy

In general, the solvation free energy is calculated using quantum mechanical methods by means of Equation (1). In this model, the cavitation free energy is not determined accurately by various theoretical models, such as the Claverie-Pierotti method [34–36]. This work therefore develops a model based on a structure–property relationship using the experimental solvation free energies of 42 liquid organic compounds, to allow for more accurate calculation of the cavitation free energy. Table 3 shows the electrostatic, dispersive, and repulsive free energies calculated using the C-PCM and Amovilli-Mennucci methods for this group of 42 organic molecules. Using the Amovilli-Mennucci approximation, the free energy for cavity construction, ΔG_{cav} , was calculated using Equation (28):

$$\Delta G_{cav} = \Delta G_{solv(exp)} - \Delta G_{elec} - \Delta G_{rep} - \Delta G_{disp} \quad (28)$$

The term $\Delta G_{solv(exp)}$ corresponds to the experimental solvation free energies reported for these organic molecules in 1-octanol and water. Equation (28) was used by Amovilli and Mennucci to calculate the cavitation free energy: they plotted the free energy associated with cavity construction as a function of the cavity areas (A_{cavity}) of the molecules, obtaining a linear relationship of the form:

$$\Delta G_{cav} = \beta \times A_{cavity} + \alpha \quad (29)$$

In Equation (29), the terms α and β are parameters obtained by linear regression. This model has been proposed in other solvation studies [51,52]. Here, we propose an equation to calculate the free energy for cavity construction based on molecular descriptors calculated with the C-PCM and GEPOL-GB methods included in the GAMESS-US package. The linear regression model is shown in Equation (30):

$$\Delta G_{cav(estimate)} = \sum_{i=1}^n \sigma_i A_i + \sum_{j=1}^m k_j X + b \quad (30)$$

where $n = 4$ and $m = 9$. The term i corresponds to the different atoms in the solute molecule ($i = \text{C, H, N, and O}$). The factor σ_i is associated with each atom, and k_j is a factor related to each functional group present in the molecule. This equation, obtained by linear regression, includes the surface areas of the atoms in the solute molecule and a factor associated with the functional groups present in the molecule. The term X is equal to one if the functional group is present and zero otherwise.

The parameter b is a constant term obtained by linear regression, and A_i is the surface area of atom i obtained using the GEPOL-GB method [46]. Models similar to Equations (29) and (30), based on statistical mechanics, have been reported by Marenich, Cramer, Truhlar, and Kollman [53–55].

The parameter values obtained by linear regression to estimate the free energy required for cavity construction are given in Table 4. This model was parameterized using the root mean square error ($RMSE$), defined by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n |\Delta G_{cav} - \Delta G_{cav(estimate)}|^2}{n}} \quad (31)$$

Table 4: σ_i and k_j coefficients obtained by multilinear regression for the estimation of the cavitation free energy in 1-octanol and water.

k_j	1-Octanol kJ mol ⁻¹	Water kJ mol ⁻¹	σ_i	1-Octanol kJ mol ⁻¹ Å ⁻²	Water kJ mol ⁻¹ Å ⁻²
$k_1 = k_{aromatics}$	-10.45	-33.88	σ_{carbon}	0.44	1.08
$k_2 = k_{diol}$	0.20	7.83	σ_{oxygen}	0.12	-0.40
$k_3 = k_{ethers}$	2.43	-1.40	$\sigma_{hydrogen}$	0.07	-0.03
$k_4 = k_{ketones}$	8.04	1.22	$\sigma_{nitrogen}$	0.17	2.36
$k_5 = k_{cycloalkane}$	-2.64	-1.70			
$k_6 = k_{esters}$	9.67	8.75			
$k_7 = k_{aldehydes}$	3.12	2.29			
$k_8 = k_{amines}$	-2.56	-32.33			
$k_9 = k_{acids}$	-----	6.96			
b	9.02	3.84			
$RMSE$	0.67	1.25			

The parameterization was performed for water and 1-octanol, with coefficients of determination of $R^2 = 0.990$ and $R^2 = 0.986$, respectively. In the case of 1-octanol, the coefficient k_j was not determined for the acid functional group.

As presented in Table 5, the solvation free energies in water and 1-octanol were obtained using the free energy of cavity formation, as estimated by Equation (30). The results are in good agreement for the 42 molecules included in the parameterization process.

As illustrated in Figures 2 and 3, the estimated solvation free energy values are plotted against the experimental solvation free energy in water and 1-octanol. For the multivariate model, the linear regression curve demonstrates a strong correlation for the solvation free energies in both water and 1-octanol. The errors are generally larger for 1-octanol solvation energies than for water solvation energies, as evident from the RMSE values in Table 4: RMSE = 1.25 kJ mol⁻¹ for water and RMSE = 0.67 kJ mol⁻¹ for 1-octanol. Although the RMSE is smaller for

1-octanol, the relative errors (see values in Table 5) show that the deviations for 1-octanol are more systematic and depend on the functional group. This indicates that the use of macroscopic solvent parameters for 1-octanol introduces systematic errors, because this solvent exhibits a complex molecular structure that is not adequately represented by a uniform dielectric continuum. 1-Octanol

forms dynamic aggregates through hydrogen bonding and hydrophobic interactions, creating a heterogeneous environment that affects solute partitioning in ways that simple continuum models cannot capture. This is particularly evident for molecules with polar functional groups (e.g., alcohols, ketones), where specific interactions with the 1-octanol hydroxyl group are important.

Table 5: Estimated free energies of solvation in water and 1-octanol for the 42 molecules studied.

	1-Octanol			Water		
	ΔG_{exp} kJ mol ⁻¹	$\Delta G(\text{Estimate})$ kJ mol ⁻¹	Relative Error	ΔG_{exp} kJ mol ⁻¹	$\Delta G(\text{Estimate})$ kJ mol ⁻¹	Relative Error
acetone	-13.18	-14.62	0.11	-16.11	-17.14	0.07
2-butanone	-15.82	-15.42	0.03	-15.23	-15.16	0.00
2-pentanone	-18.20	-18.26	0.00	-14.77	-14.67	0.01
3-pentanone	-18.24	-16.90	0.07	-14.27	-12.86	0.10
2-hexanone	-21.00	-21.24	0.01	-13.77	-14.32	0.03
1-propanal	-17.28	-16.98	0.02	-14.39	-14.25	0.01
1-butanal	-19.33	-19.63	0.02	-13.31	-13.45	0.01
acetic acid	-26.57	-27.04	0.02	-28.03	-28.68	0.03
propanoic acid	-28.70	-28.12	0.02	-27.07	-27.65	0.02
butanoic acid	-31.72	-31.58	0.00	-26.61	-26.38	0.01
pentanoic acid	-34.39	-34.11	0.01	-25.77	-25.51	0.01
hexanoic acid	-36.90	-37.10	0.01	-25.98	-25.26	0.03
benzene	-15.57	-16.22	0.04	-3.64	-7.35	1.03
toluene	-19.04	-18.73	0.02	-3.72	-2.19	0.41
ethylbenzene	-21.26	-20.91	0.01	-3.35	-1.16	0.67
methanol	-15.82	-15.70	0.02	-21.38	-20.55	0.02
ethanol	-18.24	-18.41	0.00	-20.96	-19.48	0.05
1-propanol	-21.00	-21.01	0.01	-20.21	-18.52	0.07
1-butanol	-23.89	-24.14	0.01	-19.75	-21.69	0.11
1-pentanol	-26.78	-26.40	0.02	-18.70	-18.79	0.01
1-hexanol	-29.54	-30.21	0.02	-18.24	-17.90	0.01
1-heptanol	-32.43	-33.21	0.02	-17.74	-17.54	0.01
2-propanol	-19.33	-19.96	0.03	-19.92	-17.95	0.10
1,2-ethanediol	-31.13	-31.13	0.00	-38.91	-38.90	0.01
dimethyl ether	-8.62	-13.40	0.01	-8.03	-6.05	0.14
diethyl ether	-12.09	-13.93	0.11	-7.36	-7.23	0.18
propyl methyl ether	-15.19	-14.27	0.08	-6.95	-14.93	0.04
methyl acetate	-14.81	-17.52	0.04	-13.89	-14.27	0.08
ethyl acetate	-16.99	-16.04	0.03	-12.97	-13.13	0.10
methyl propionate	-16.99	-20.16	0.06	-12.26	-8.64	0.07
methyl butanoate	-19.21	-11.01	0.05	-11.84	-7.14	0.28
cyclopentane	-11.09	-14.15	0.01	5.02	4.86	0.41
cyclohexane	-14.48	-16.64	0.02	5.15	5.32	0.06
methylcyclohexane	-16.23	-9.46	0.03	7.15	9.62	0.25
n-pentane	-10.25	-12.48	0.08	9.75	9.76	0.02
n-hexane	-12.59	-15.40	0.01	10.42	10.32	0.07
n-heptane	-15.65	-18.06	0.01	10.96	11.00	0.06
n-octane	-17.49	-21.31	0.04	12.09	11.04	0.49
n-propylamine	-19.96	-21.85	0.08	-17.53	-17.41	0.03
n-butylamine	-22.30	-18.52	0.02	-17.95	-17.03	0.03
diethylamine	-19.87	-14.62	0.05	-17.03	-17.14	0.02
dipropylamine	-25.19	-15.42	0.02	-15.31	-15.16	0.02

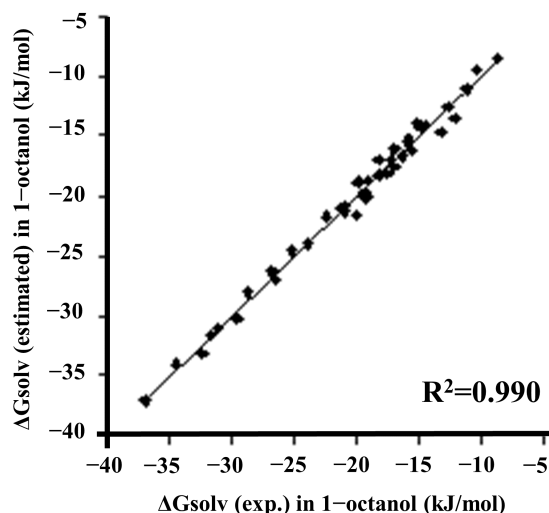


Figure 2: Solvation free energy estimated in 1-octanol versus the experimental solvation free energy.

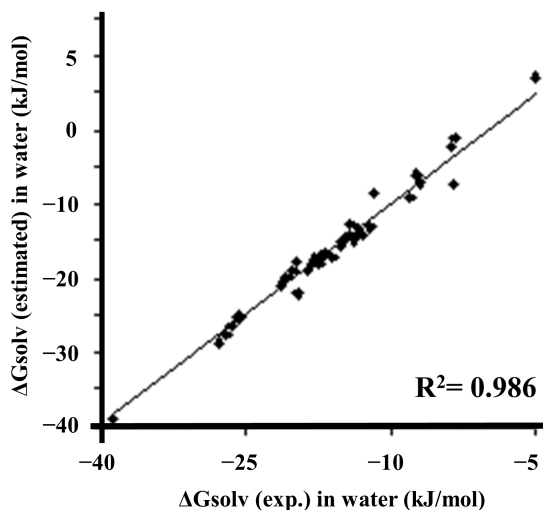


Figure 3: Solvation free energy estimated in water versus the experimental solvation free energy.

The lowest correlation was obtained for the solvation free energy in water. This is due to the 6-31G(2p,2d) basis set used to estimate the electrostatic contribution and the minimal basis set used in the Amovilli-Mennucci method to determine the dispersion and repulsion free energies. In other studies, these solvation free energies have been estimated using additional molecular descriptors, such as hydrogen-bond acceptor and donor indices, which has improved the estimation of these properties [56,57].

4.2. Prediction of the Partition Coefficient $K_{o/w}$

Using the estimated solvation free energies and Equation (27), the partition coefficient was determined. Table 6

shows the estimated partition coefficient values, which vary considerably from the reported experimental values.

The coefficient of determination obtained was $R^2 = 0.974$, using the solvation free energy values calculated in this work (see correlation in Figure 4). These results show a better correlation than those reported by Amovilli-Mennucci using their model to estimate the cavitation free energy.

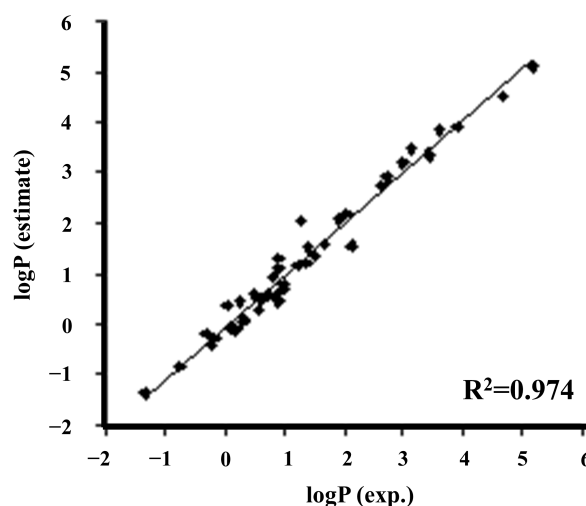


Figure 4: Partition coefficients estimated using the solvation free energy in 1-octanol and water.

At the same time, Table 6 presents the partition coefficients with their respective relative errors, showing that the greatest estimation error occurred for 2-propanol and molecules containing the carbonyl functional group. Additionally, the amines exhibit high relative errors in the estimated partition coefficients. Nevertheless, the results obtained for the partition coefficient using solvation free energies estimated with the cavitation free energy model are better than the values obtained by Amovilli-Mennucci. It is also important to note that the distribution of a solute between two immiscible solvents is fundamentally a transfer process; therefore, using solvation free energies is not entirely suitable for estimating partition coefficients, a limitation that should be addressed using a model that determines the transfer free energy directly. The 6-31G(2p,2d) basis set used to calculate the electrostatic contribution to the solvation free energy remains appropriate; however, a systematic source of error is the minimal basis set used by the Amovilli-Mennucci method to calculate the dispersion and repulsion free energies.

Table 6: Estimated partition coefficients using the solvation free energies in water and 1-octanol.

	logP Estimate	logP Exp.	Relative Error		logP Estimate	logP Exp.	Relative Error
acetone	-0.44	-0.24	0.84	1-heptanol	2.75	2.62	0.05
2-butanone	0.05	0.29	0.84	2-propanol	0.36	0.05	6.14
2-pentanone	0.63	0.91	0.31	1,2-ethanediol	-1.36	-1.36	0.00
3-pentanone	0.71	0.99	0.28	dimethyl ether	-0.09	0.10	1.86
2-hexanone	1.21	1.38	0.12	diethyl ether	1.29	0.89	0.45
1-propanal	0.48	0.59	0.19	propyl methyl ether	1.17	1.21	0.03
1-butanal	1.08	0.88	0.23	methyl acetate	-0.11	0.18	1.64
acetic acid	-0.29	-0.17	0.68	ethyl acetate	0.57	0.73	0.22
propanoic acid	0.08	0.33	0.75	methyl propionate	0.51	0.82	0.38
butanoic acid	0.91	0.79	0.15	methyl butanoate	2.02	1.29	0.56
pentanoic acid	1.51	1.39	0.08	cyclopentane	3.18	3.00	0.06
hexanoic acid	2.08	1.92	0.08	cyclohexane	3.33	3.44	0.03
benzene	1.55	2.13	0.27	methylcyclohexane	3.85	3.61	0.07
toluene	2.90	2.73	0.06	n-pentane	3.34	3.39	0.01
ethylbenzene	3.46	3.15	0.10	n-hexane	3.90	3.90	0.00
methanol	-0.85	-0.77	0.10	n-heptane	4.51	4.66	0.03
ethanol	-0.19	-0.31	0.40	n-octane	5.09	5.18	0.02
1-propanol	0.44	0.25	0.74	n-propylamine	0.57	0.48	0.18
1-butanol	0.43	0.88	0.51	n-butylamine	0.78	0.97	0.20
1-pentanol	1.33	1.51	0.12	diethylamine	0.26	0.58	0.55
1-hexanol	2.16	2.03	0.06	dipropylamine	1.57	1.67	0.06

As shown in Table 6, molecules with significant van der Waals contributions to their solvation in 1-octanol, such as 2-propanol and dimethyl ether, exhibit the largest deviations. In these cases, the minimal STO-3G basis set is too limited to describe the anisotropic electron density distributions necessary for accurate dispersion interactions. This limitation is particularly pronounced for molecules containing secondary alcohol groups, in which the steric environment surrounding the hydroxyl group modulates dispersion interactions with the solvent in a manner that requires a more flexible basis set. In future work, basis sets of at least triple-zeta quality with polarization functions, such as 6-311+G(d,p), will be employed to improve the accuracy of these non-electrostatic contributions. It is important to note that the deviations observed in Table 6 are not random but reflect a fundamental limitation of the approach based on the difference in solvation energies (see Equation (27)). This method assumes that the transfer free energy between phases is exactly equal to the difference in solvation energies, a simplification that ignores activity effects, conformational changes, and molecular association in the organic phase. Additionally, although the GEPOL-GB method for cavity construction is computationally efficient, it relies on fixed van der Waals radii and a single molecular conformation optimized in vacuum. Consequently, this approach does not account for conformational flexibility, which is particularly problematic for molecules capable of adopting multiple low-energy conformations with different solvent-

accessible surface areas. For example, 2-propanol can rotate around its C–O bond, altering the exposure of its hydroxyl group to the solvent. 1,2-Ethanediol, by contrast, shows an excellent prediction because its conformation is relatively constrained by intramolecular hydrogen bonding, whereas flexible molecules like 2-propanol and methyl butanoate can adopt multiple conformations, and using a single optimized structure fails to capture the conformational ensemble that contributes to solvation.

In this paper, we also propose two quantitative structure–property relationship (QSPR) models for estimating the partition coefficient using properties obtained with the C-PCM method. These QSPR models incorporate the electrostatic free energy calculated in 1-octanol and water, the atomic surfaces of the atoms composing the molecules (determined using the GEPOL-GB method), and factors associated with the functional groups present in the molecule. In this case, 50 molecules were used for parameterization. The first QSPR model, defined as Model A, is shown in Equation (32):

$$K_{o/w} = \log P = a \times (\Delta G_{elec(water)} - \Delta G_{elec(1-octanol)}) + \sum_{i=1}^n \sigma_i A_i + \sum_{j=1}^m k_j X + b \quad (32)$$

where $n = 4$ and $m = 9$. The term i corresponds to the different atom types in the solute molecule ($i = \text{C, H, N and O}$). The factor σ_i is the atomic tensor associated with each atom and k_j is a factor related to each functional

group present in the molecule. The terms $\Delta G_{elec(water)}$ and $\Delta G_{elec(1-octanol)}$ correspond to the electrostatic free energies of solvation and the term X is equivalent to one if the functional group is present and zero otherwise.

Model A achieves a correlation coefficient of $R^2 = 0.991$, making it simpler than the model of Equation (30), which requires greater computational cost for estimating the free energy of dispersion and repulsion. This model is also more accessible and better suited to large molecules.

All parameters of Equation (32) were obtained using the root mean square error ($RMSE$), defined by Equation (33):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n |\log P_{(exp)} - \log P_{(estimate)}|^2}{n}} \quad (33)$$

Table 7 contains the values obtained by multivariate regression of the coefficients associated with Model A. Unlike the model that uses the cavitation free energy, Model A directly yields the distribution coefficient. The RMSE was 0.16, and the statistical analysis shows that the most influential factors are the presence of aromatic and amine functional groups. Model A is parameterized directly using experimental logP values, which incorporates the transfer free energy between phases and avoids the systematic errors inherent in the solvation energy difference approach. The chemical significance of the functional group coefficients in Model A reflects the distinct solvation environments in water versus 1-octanol: negative coefficients for aromatic groups (-3.168) and amines (-3.789) indicate their strong preference for the organic phase due to π - π stacking and hydrophobic effects, while the positive coefficient for acids (0.189) reflects their greater affinity for water through hydrogen bonding. The ketone coefficient (-1.457) reflects the balance between carbonyl hydrogen bonding with water and dispersion interactions with the alkyl chain of 1-octanol,

whereas the ether (-1.208) and ester (-0.783) coefficients reflect their intermediate polarity and their ability to act as hydrogen bond acceptors in both phases. The atomic surface coefficients enhance the performance of Model A by accounting for molecular size and shape effects, with the positive carbon surface coefficient (0.086) capturing hydrophobic contributions and the negative oxygen surface coefficient (-0.063) reflecting polar solvation effects. This combination of direct transfer free energy modeling, empirically calibrated functional group contributions, and atomic surface descriptors allows Model A to achieve superior correlation ($R^2 = 0.991$, $RMSE = 0.16$) while maintaining computational simplicity and chemical interpretability. **Table 8** shows the estimated partition coefficient values and their absolute relative errors using Model A, where dimethyl ether and 2-propanol show the greatest relative error. **Figure 5** shows the estimated partition coefficient values for Model A relative to the experimental partition coefficient.

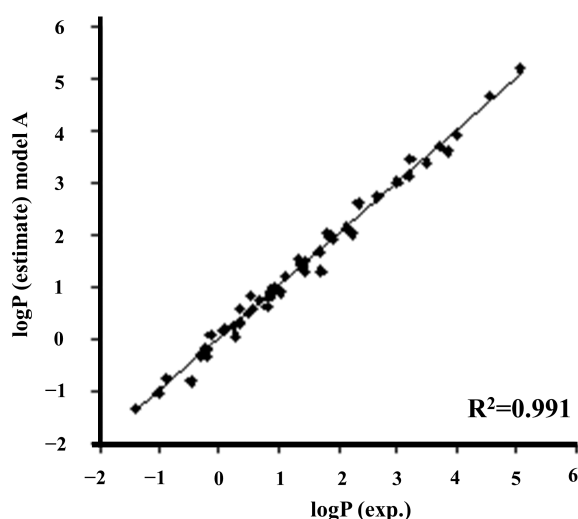


Figure 5: Partition coefficients estimated using the model A.

Table 7: σ_i and k_j coefficients obtained by multilinear regression for the estimation of the partition coefficients using the model A.

σ_i	$\dot{A}^{-2}$	k_j	
σ_{carbon}	0.0864	$k_{ketones}$	-1.4573
σ_{oxygen}	-0.0631	$k_{aldehydes}$	-0.9049
$\sigma_{nitrogen}$	0.1891	k_{acids}	0.1894
$\sigma_{hydrogen}$	-0.0028	$k_{aromatics}$	-3.1680
		$k_{cycloalkanes}$	-0.0152
Coefficient	mol/kJ	k_{ethers}	-1.2084
a	0.3286	k_{esters}	-0.7834
Coefficient		k_{amines}	-3.7888
b	-0.0797	$k_{alcohols}$	-0.9449

Table 8: Estimated partition coefficients using the model A.

	logP Exp.	logP Estimate	Relative Error		logP Exp.	logP Estimate	Relative Error
acetone	-0.24	-0.24	0.01	2-propanol	0.05	0.25	4.09
2-butanone	0.29	0.35	0.19	2-butanol	0.61	0.81	0.33
2-pentanone	0.91	0.88	0.04	1,2-ethanediol	-1.36	-1.39	0.02
3-pentanone	0.99	0.93	0.06	1,3-propanediol	-1.04	-1.01	0.03
2-hexanone	1.38	1.41	0.02	1,4-butanediol	-0.83	-0.48	0.42
ethanal	-0.34	-0.21	0.39	dimethyl ether	0.10	-0.15	2.47
n-propanal	0.59	0.34	0.43	diethyl ether	0.89	1.02	0.15
n-butanal	0.88	0.88	0.00	ether dipropyl	2.03	2.24	0.10
n-pentanal	1.31	1.43	0.09	methyl propyl ether	1.21	1.11	0.08
acetic acid	-0.17	-0.23	0.37	methyl acetate	0.18	0.09	0.51
propanoic acid	0.33	0.35	0.06	ethyl acetate	0.73	0.67	0.08
butanoic acid	0.79	0.85	0.08	methyl propanoate	0.82	0.55	0.33
pentanoic acid	1.39	1.38	0.01	methyl butanoate	1.29	1.71	0.33
hexanoic acid	1.92	1.91	0.00	cyclopentane	3.00	2.99	0.00
benzene	2.13	2.14	0.00	cyclohexane	3.44	3.23	0.06
toluene	2.73	2.66	0.03	methylcyclohexane	3.61	3.83	0.06
ethylbenzene	3.15	3.19	0.01	n-pentane	3.39	3.48	0.03
propylbenzene	3.69	3.72	0.01	n-hexane	3.90	4.01	0.03
methanol	-0.77	-0.87	0.13	n-heptane	4.66	4.56	0.02
ethanol	-0.31	-0.30	0.02	n-octane	5.18	5.08	0.02
1-propanol	0.25	0.25	0.01	n-propylamine	0.48	0.47	0.01
1-butanol	0.88	0.87	0.01	n-butylamine	0.97	1.01	0.04
1-pentanol	1.51	1.33	0.12	n-pentylamine	1.49	1.45	0.02
1-hexanol	2.03	1.82	0.10	diethylamine	0.58	0.57	0.01
1-heptanol	2.62	2.35	0.10	dipropylamine	1.67	1.68	0.00

Finally, we present a second model to determine the partition coefficient of the 50 organic molecules, defined as Model B. Model B uses properties such as the non-bonding free energy, given by Equation (34):

$$\Delta G_{non-bond} = \Delta G_{elec} + \Delta G_{disp} + \Delta G_{rep} \quad (34)$$

This multivariate regression model includes the atomic surfaces A_i of the atoms composing the molecules and the coefficients associated with their functional groups. Model B is shown in Equation (35):

$$\log P = a \times \Delta G_{non-bond(1-octanol)} + b \times \Delta G_{non-bond(water)} + \sum_{i=1}^n \sigma_i A_i + \sum_{j=1}^n k_j X_j + c \quad (35)$$

Similarly, the coefficients in Equation (35) were obtained using the experimental partition coefficient values in the multivariate regression. **Figure 6** shows the estimated values for Model B versus the experimental partition coefficient values. Model B achieves a correlation coefficient of 0.986 and a Root Mean Square Error (RMSE) of 0.17 relative to the experimental data. **Table 9** shows the parameter values obtained for Model B.

In this model, no coefficients are associated with the aldehyde, alcohol, ester, and cycloalkane functional groups, which makes it a simpler model, although determining coefficients for these functional groups would have improved the multivariate correlation. However, Model B has a higher computational cost because it involves the calculation of dispersive and repulsive free energies. The systematic deviations observed for alcohols (2-propanol, relative error 6.14), amines (diethylamine, relative error 0.55), and carbonyl-containing compounds (acetone, relative error 0.84; methyl acetate, relative error 1.64) indicate that hydrogen bonding and specific intermolecular interactions are not fully captured by the current models, reflecting the inherent limitations of continuum solvation methods that represent solvents as uniform dielectric media. Including additional descriptors, such as hydrogen-bond donor/acceptor ability, molecular dipole moment, or polar surface area, would likely improve predictive capability by explicitly encoding information about the solute's capacity for specific interactions. While the atomic surface coefficients in these models partially capture polar solvation effects, they do not distinguish between hydrogen bond donors and acceptors, nor do they account for interaction strength or directionality.

Table 9: σ_i and k_j coefficients obtained by multilinear regression for estimating the partition coefficients using the model B.

σ_i	A^{-2}	k_j	
σ_{carbon}	0.1067	$k_{ketones}$	-0.6333
σ_{oxygen}	-0.0347	$k_{aldehyds}$	-----
$\sigma_{nitrogen}$	0.2978	k_{acids}	1.0984
$\sigma_{hydrogen}$	0.0130	$k_{aromatics}$	-2.7048
coefficients	$\frac{mol}{kJ}$	$k_{cycloalkanes}$	-----
		k_{ethers}	-0.8672
a	-0.0341	k_{esters}	-----
b	0.2667	k_{amines}	-4.0184
coefficient		$k_{alcohols}$	-----
c	-0.2670		

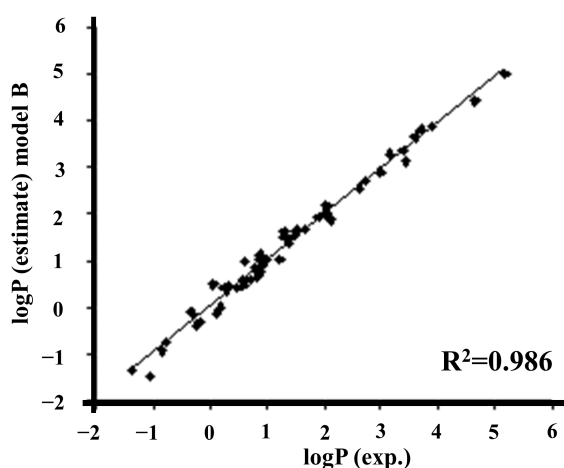


Figure 6: Partition coefficients estimated using model B.

Additionally, Table 10 presents the logP values predicted by Model B for 50 organic molecules; the largest relative errors were observed for molecules containing ester, hydroxyl, and aldehyde functional groups, as the corresponding parameters were not included in the analysis. Furthermore, both QSPR models are limited to molecules containing a single functional group within their structures.

The linear regression models assume additive contributions from atomic surfaces and functional groups. However, the results indicate that this additivity assumption breaks down for molecules with intermediate alkyl chain lengths (e.g., 2-butanone), where the balance between hydrophilic and hydrophobic contributions is most sensitive to the local molecular environment. This suggests that non-linear effects, such as synergistic interactions between the carbonyl group and the alkyl chain, are not fully captured by the current model. Future studies should explore non-linear regression techniques, such as random forests or neural networks, using a larger and more di-

verse training set to better capture these complex structure-property relationships. Furthermore, $K_{o/w}$ describes the equilibrium distribution of a solute between two immiscible liquid phases and is fundamentally a property of the transfer process between solvents, not a property of solvation from vacuum. While solvation free energies (ΔG_{solv}) describe the transfer from vacuum to a single solvent, the transfer free energy ($\Delta G_{transfer}$) describes the transfer between solvents and is directly related to logP. Using solvation free energies to estimate logP (Equation (27)) introduces systematic errors because it assumes that the difference in solvation free energies (both referenced to vacuum) equals the transfer free energy between phases. This assumption neglects the differences in activity coefficients between solvents, conformational changes induced by different solvent environments, and specific solute-solvent and solvent-solvent interactions that are not captured by continuum models. These limitations explain why the solvation energy difference approach in Table 6 shows larger deviations ($R^2 = 0.974$) compared with Model A ($R^2 = 0.991$), which is parameterized directly against the experimental logP values and effectively models the transfer free energy.

Additionally, the dispersive and repulsive free energies in Model B account for van der Waals interactions and Pauli exclusion effects that are particularly significant in 1-octanol, where the long alkyl chain provides extensive dispersion interactions with hydrophobic regions of the solute, while repulsive terms capture excluded-volume effects governing molecular packing and cavity formation in both phases. However, these contributions are calculated using a minimal basis set (STO-3G) in the Amovilli-Mennucci method, which systematically underestimates dispersion interactions for molecules with diffuse electron densities and fails to capture effects that are crucial for accurate logP prediction, limiting their overall contribution to predictive accuracy. Including these

terms in Model B yields only a marginal improvement in correlation ($R^2 = 0.986$) compared with Model A ($R^2 = 0.991$), while introducing additional computational cost and potential error propagation—demonstrating that the dominant factors for logP prediction are electrostatic interactions and molecular surface descriptors, rather than explicit dispersion-repulsion calculations. The systematic errors in the dispersion-repulsion terms also explain why the solvation energy difference approach (see values in Table 6) exhibits larger deviations for molecules

in which van der Waals interactions dominate solvation in 1-octanol, such as alkanes and ethers, whereas Model A effectively circumvents this limitation through direct parameterization against experimental data. Consequently, despite their theoretical importance, the practical contribution of dispersive and repulsive free energies to predictive accuracy is limited by current methodological constraints, making Model A the most computationally efficient and accurate choice for routine logP prediction.

Table 10: Estimated partition coefficients using the model B.

	logP Exp.	logP Estimate	Relative Error		logP Exp.	logP Estimate	Relative Error
acetone	−0.24	−0.39	0.64	1,2-ethanediol	−1.36	−1.34	0.01
2-butanone	0.29	0.36	0.23	1,3-propanediol	−1.04	−1.50	0.44
2-pentanone	0.91	0.90	0.01	1,4-butanediol	−0.83	−0.93	0.13
3-pentanone	0.99	1.02	0.03	2-propanol	0.05	0.48	8.64
2-hexanone	1.38	1.45	0.05	2-butanol	0.61	0.96	0.57
ethanal	−0.34	−0.11	0.68	dimethyl ether	0.10	−0.12	2.21
n-propanal	0.59	0.46	0.22	diethyl ether	0.89	1.13	0.27
n-butanal	0.88	1.03	0.17	ether dipropyl	2.03	2.18	0.07
n-pentanal	1.31	1.61	0.23	methyl propyl ether	1.21	1.04	0.14
acetic acid	−0.17	−0.34	0.98	methyl acetate	0.18	0.01	0.94
propanoic acid	0.33	0.44	0.34	ethyl acetate	0.73	0.59	0.19
butanoic acid	0.79	0.83	0.05	methyl propanoate	0.82	0.63	0.23
pentanoic acid	1.39	1.39	0.00	methyl butanoate	1.29	1.53	0.19
hexanoic acid	1.92	1.93	0.00	cyclopentane	3.00	2.90	0.03
benzene	2.13	1.88	0.12	cyclohexane	3.44	3.13	0.09
toluene	2.73	2.72	0.00	methylcyclohexane	3.61	3.64	0.01
ethylbenzene	3.15	3.29	0.04	n-pentane	3.39	3.35	0.01
propylbenzene	3.69	3.81	0.03	n-hexane	3.90	3.88	0.01
methanol	−0.77	−0.74	0.04	n-heptane	4.66	4.44	0.05
ethanol	−0.31	−0.14	0.54	n-octane	5.18	4.99	0.04
1-propanol	0.25	0.43	0.73	n-propylamine	0.48	0.42	0.13
1-butanol	0.88	0.74	0.16	n-butylamine	0.97	0.98	0.01
1-pentanol	1.51	1.66	0.10	n-pentylamine	1.49	1.54	0.03
1-hexanol	2.03	2.02	0.00	diethylamine	0.58	0.58	0.00
1-heptanol	2.62	2.56	0.02	dipropylamine	1.67	1.67	0.00

4.3. Validation of QSPR Models with Additional Molecules

The results presented in Table 11 demonstrate the predictive performance of Models A and B for ten additional organic molecules that were not included in the training set. These molecules encompass a chemically diverse range of functional groups, including secondary alcohols (2-pentanol), aldehydes (n-hexanal), amines (hexylamine), ethers (dibutyl ether, dioxane, and anisole), esters (propyl acetate), aromatics (n-butylbenzene), and polyols (glycerol). Both models show generally good agreement with the experimental logP values, with Model A achieving a

correlation coefficient of $R^2 = 0.991$ and Model B achieving $R^2 = 0.986$, confirming the robustness and general applicability of the QSPR approach across different chemical families. For most compounds, the predicted logP values are within 0.2–0.3 log units of the experimental values, with n-butylbenzene showing nearly perfect prediction by both models (Model A: 4.26, Model B: 4.38, Exp: 4.38) and glycerol showing the largest deviations (Model A: −2.02, Model B: −2.59, Exp: −1.76), attributable to the strong hydrogen bonding capabilities of the triol, which are not fully captured by continuum solvation models.

Table 11: Calculated electrostatic, dispersion, and repulsion free energies in 1-octanol and water, and predicted partition coefficients (logP) using Models A and B for ten additional organic molecules, compared with experimental values.

	1-Octanol			Water			logP (Model A)	logP (Model B)	logP (Exp)
	ΔG_{elec} (kJ mol ⁻¹)	ΔG_{disp} (kJ mol ⁻¹)	ΔG_{rep} (kJ mol ⁻¹)	ΔG_{elec} (kJ mol ⁻¹)	ΔG_{disp} (kJ mol ⁻¹)	ΔG_{rep} (kJ mol ⁻¹)			
2-pentanol	-11.05	-61.63	15.57	-12.43	-55.73	19.41	1.35	1.66	1.19
n-hexanal	-13.68	-68.54	16.74	-15.61	-61.97	20.92	1.96	2.17	1.78
hexylamine	-10.29	-73.43	19.37	-11.55	-66.40	24.18	4.98	2.08	2.06
dibutyl ether	-6.99	-90.04	23.68	-7.91	-81.42	29.54	3.27	4.01	3.21
propyl acetate	-14.90	-65.19	14.31	-16.82	-58.96	17.87	1.12	1.14	1.24
n-butylbenzene	-5.44	-85.65	21.13	-6.07	-77.45	26.40	4.26	4.38	4.38
dioxane	-13.10	-53.26	11.21	-14.69	-48.16	13.98	-0.14	-0.67	-0.27
anisole	-10.04	-64.81	14.02	-11.30	-58.62	17.49	2.03	1.88	2.11
glycerol	-27.11	-54.94	9.87	-30.34	-49.67	12.34	-2.02	-2.59	-1.76
3-pentanone	-14.31	-76.70	18.87	-16.40	-69.37	23.56	1.96	2.06	2.03

The prediction for hexylamine by Model A (4.98 versus 2.06 experimental) reveals a significant overestimation, suggesting that the amine functional group coefficient in Model A may require further refinement for primary amines with longer alkyl chains, whereas Model B provides a much more accurate prediction (2.08 versus 2.06), demonstrating the value of including dispersive and repulsive contributions for molecules where van der Waals interactions dominate solvation in the organic phase. For ethers and esters, both models perform well. Dibutyl ether was predicted to have logP values of 3.27 (Model A) and 4.01 (Model B), compared with the experimental value of 3.21, whereas propyl acetate was predicted to have logP values of 1.12 (Model A) and 1.14 (Model B), compared with the experimental value of 1.24. These results indicate that the atomic surface descriptors effectively capture the hydrophobic contributions of alkyl chains. The prediction for anisole (Model A: 2.03, Model B: 1.88, Exp: 2.11) confirms that the aromatic coefficient (-3.168 in Model A) appropriately accounts for the π - π interactions and hydrophobic character of the methoxybenzene structure. For dioxane, both models correctly predict a negative logP value (Model A: -0.14 , Model B: -0.67 , Exp: -0.27), reflecting the hydrophilic nature of the cyclic ether due to its two oxygen atoms, although Model A provides better accuracy owing to its direct parameterization against experimental data. The 3-pentanone prediction (Model A: 1.96, Model B: 2.06, Exp: 2.03) demonstrates that the ketone functional group coefficient successfully captures the balance between carbonyl hydrogen bonding and hydrophobic alkyl chain effects for symmetrical ketones. These results collectively confirm that Model A, despite its computational simplicity, provides predictions comparable to or better than Model B for most compound classes, with the exception of long-chain amines, where dispersion contributions become important—highlighting the complemen-

tary nature of both approaches across different chemical spaces.

4.4. Deviations and Limitations

The deviations observed in Table 6 between the estimated and experimental logP values are systematic rather than random and can be attributed to several factors. First, it is important to note a fundamental limitation of the solvation energy difference approach illustrated in Equation (27): this approach assumes that the free energy of transfer is exactly equal to the difference in solvation energies, an assumption that fails to account for activity effects, conformational changes, and molecular association in the organic phase. This limitation is most severe for molecules such as 2-propanol and dimethyl ether, which can adopt different conformations or interact specifically with the solvent. Second, the use of a minimal basis set for the dispersion and repulsion terms is a significant concern: the Amovilli-Mennucci method uses a minimal basis set (STO-3G) by default, which inadequately describes the anisotropic electron density distributions necessary for accurate dispersion interactions. This is particularly relevant for molecules in which van der Waals interactions predominate over other forces during solvation in 1-octanol. Third, the GEPOL-GB cavity model, which uses fixed van der Waals radii and a single vacuum-optimized conformation, does not account for conformational flexibility or averaging, presenting a significant challenge for molecules that exhibit multiple low-energy conformations. Fourth, the assumptions underlying linear regression models are relevant: the additivity assumption becomes invalid for molecules with intermediate alkyl chain lengths, owing to the sensitivity of their hydrophilic and hydrophobic contributions to the local molecular environment. Furthermore, macroscopic solvent parameters for 1-octanol fail to cap-

ture its microheterogeneous structure (monomers, dimers, aggregates), leading to systematic errors for molecules with polar functional groups.

Finally, it is important to emphasize that Model A (Equation (32)) substantially mitigates these limitations by being parameterized directly against experimental logP values, achieving a considerably better correlation ($R^2 = 0.991$, RMSE = 0.16) than the solvation energy difference approach ($R^2 = 0.974$). By contrast, the Amovilli-Mennucci method systematically underestimates dispersion interactions for molecules with diffuse electron densities and fails to capture effects that are crucial for accurate logP prediction, limiting its overall contribution to predictive accuracy; including these terms in Model B yields only a marginal improvement, with a correlation coefficient of $R^2 = 0.986$. Additionally, a limitation of the current QSPR models is their use of binary indicators for functional groups, whereby a value of 1 is assigned if the functional group is present and 0 otherwise, regardless of the number of identical groups in the molecule. For example, 1,2-ethanediol receives the same alcohol indicator as methanol or ethanol, despite having two hydroxyl groups that substantially increase its hydrogen bonding capacity and hydrophilicity. This simplification is particularly problematic for polyols such as glycerol, which contains three hydroxyl groups and shows larger deviations (Model A: -2.02 , Model B: -2.59 , Exp: -1.76) compared with monohydric alcohols, suggesting that the models underestimate the hydrophilicity of polyfunctional compounds. Using functional group counts (e.g., number of OH groups, NH groups, or carbonyl groups) rather than binary presence would allow the models to capture the additive contributions of multiple identical groups to overall partitioning behavior, improving predictions for polyfunctional molecules. Furthermore, for compounds containing multiple different functional groups simultaneously (e.g., amino alcohols, hydroxy acids, keto-esters), interaction terms between functional groups could be incorporated to account for non-additive effects, such as intramolecular hydrogen bonding, which can significantly modify solvation properties compared with the sum of individual group contributions. The binary indicator approach was chosen in this work for its simplicity and interpretability, as well as the limited number of polyfunctional compounds in the training set.

5. Conclusions

In this work, the partition coefficients of several organic molecules were estimated using quantitative structure–property relationship (QSPR) models. Constructing these models required only the electrostatic free energy ob-

tained by the C-PCM//PBE-VWN/6-31G(2p,2d) method and the dispersive and repulsive free energies calculated by the Amovilli-Mennucci method. Of the three proposed models containing parameters calculated with the C-PCM method, Model A achieved the highest coefficient of determination ($R^2 = 0.991$), with an RMSE of 0.16. This model relates the electrostatic free energy and the terms associated with the functional groups present in the molecule to the partition coefficient. Additionally, the QSPR models obtained in this work to estimate the free energy for cavity construction in water and 1-octanol were acceptable, with correlation coefficients of $R^2 = 0.990$ and $R^2 = 0.986$ for 1-octanol and water, respectively.

While a direct comparison with alternative solvation models such as SMD, COSMO-RS, or IEFPCM would further validate the robustness of these conclusions, the primary objective of this work is to demonstrate that C-PCM-derived descriptors, when combined with QSPR modeling, can achieve high predictive accuracy for logP ($R^2 = 0.991$) as a revision and extension of the original Amovilli-Mennucci methodology. The C-PCM//PBE-VWN/6-31G(2p,2d) approach was selected because it provides a favorable balance between accuracy and computational cost. Moreover, its wide availability in the GAMESS-US package ensures accessibility for the broader research community. Although SMD and COSMO-RS incorporate more refined parameterization of non-electrostatic contributions, the excellent correlation achieved by our model suggests that the electrostatic and surface descriptors derived from C-PCM capture the essential physics governing the partitioning process. Future work will explore the application of SMD, COSMO-RS, and IEFPCM to further improve the predictive capability of these QSPR models and evaluate their general applicability across different solvation methodologies. Finally, the predictive accuracy of the proposed models would be substantially enhanced by expanding the training dataset to include a wider range of chemical functionalities, particularly molecules with multiple functional groups, heterocycles, and halogenated compounds. Further external validation using larger independent datasets not included in the parameterization process would strengthen the reliability and generalizability of the models.

List of Abbreviations

C	Carbon
COSMO	Conductor-like Screening Model
COSMO-RS	Conductor-like Screening Model for Real Solvents

C-PCM	Conductor-like Polarizable Continuum Model
DFT	Density Functional Theory
GAMESS-US	General Atomic and Molecular Electronic Structure System (US version)
GEPOL-GB	GEneration of POLyhedral cavities–Gauss-Bonnet method
H	Hydrogen
IEFPCM	Integral Equation Formalism Polarizable Continuum Model
Ko/w	Partition coefficient oil/water
logP	Logarithm of the partition coefficient
N	Nitrogen
O	Oxygen
PBE	Perdew–Burke–Ernzerhof (exchange–correlation functional)
PBE-VWN	Perdew–Burke–Ernzerhof/Vosko–Wilk–Nusair functional
PCM	Polarizable Continuum Model
QSPR	Quantitative Structure–Property Relationship
r	Linear correlation coefficient
R ²	Coefficient of determination
RHF	Restricted Hartree-Fock
RMSE	Root Mean Square Error
SCF	Self-Consistent Field
SMD	Solvation Model based on Density
STO-3G	Slater-Type Orbital approximated by 3 Gaussians
UHF	Unrestricted Hartree-Fock
vdW	van der Waals
VWN	Vosko–Wilk–Nusair (correlation functional)
wxMacMolPlt	Molecular visualization and input preparation software
ΔG _{cav}	Cavitation free energy
ΔG _{disp}	Dispersion free energy
ΔG _{elec}	Electrostatic free energy of solvation
ΔG _{rep}	Repulsion free energy
ΔG _{solv}	Solvation free energy
ΔG _{vap}	Vaporization free energy

Author Contributions

The author is solely responsible for conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, and visualization of the article. The author has read and agreed to the final version of the manuscript.

Availability of Data and Materials

The data supporting the findings of this study are available within the article and from the author upon reasonable request.

Conflicts of Interest

The author declares no conflicts of interest.

Funding

The study did not receive any external funding and was conducted using only institutional resources.

Acknowledgments

The author gratefully acknowledges the Universidad de Carabobo, Facultad Experimental de Ciencias y Tecnología, and the Laboratorio de Simulación Molecular para la Química Experimental (SIMOLQUIMEX) for providing the computational resources and institutional support necessary to carry out this work.

AI Declaration

In accordance with COPE guidance on "Authorship and AI Tools" (2023), DeepSeek (DeepSeek-V3/DeepSeek-R1) was used solely to edit existing text for language polishing, grammatical correction, stylistic refinement, and improved clarity, readability, and coherence. It was not used as an author, nor was it used to generate scientific content, data, methodology, interpretations, or conclusions. All AI-assisted edits were performed under my supervision and were critically reviewed, verified, and approved before incorporation. As the sole author, I take full responsibility for the content, accuracy, integrity, and originality of the work, including all data, calculations, interpretations, and conclusions.

References

- [1] Reinhard, M.; Drefahl, A. *Handbook for Estimating Physicochemical Properties of Organic Compounds*; John Wiley & Sons, Inc.: New York, NY, USA, 1998; Available online: <https://uat.store.wiley.com/en-us/shop/general-chemistry/handbook-for-estimating-physicochemical-properties-of-organic-compounds-p-9780471172642> (accessed on 1 February 2026).
- [2] Xiao, Z.J.; Chen, J.W.; Wang, Y.; Wang, Z.Y. In Silico Package Models for Deriving Values of Solute Parameters in Linear Solvation Energy Relationships. *SAR QSAR Environ. Res.* **2023**, *34*, 21–37. [CrossRef] [PubMed]
- [3] Wang, S.; Lin, S.-T.; Chang, J.; Goddard, W.A., III; Sandler, S.I. Application of the COSMO–SAC–BP Solvation Model to Predictions of Normal Boiling Temperatures for Environmentally Significant Sub-

- stances. *Ind. Eng. Chem. Res.* **2006**, *45*, 5426–5434. [CrossRef]
- [4] Ulrich, N.; Ebert, A. Can Deep Learning Algorithms Enhance the Prediction of Solute Descriptors for Linear Solvation Energy Relationship Approaches? *Fluid Phase Equilibria* **2022**, *555*, 113349. [CrossRef]
- [5] Zhurko, G.A.; Fedorova, A.A. QSPR Prediction of the Acidities of Carboxylic Acids and Phenols with Different Approaches. *Mol. Phys.* **2025**, *123*, e2396535. [CrossRef]
- [6] Roy, D.; Patel, C. Revisiting the Use of Quantum Chemical Calculations in LogP_{octanol-water} Prediction. *Molecules* **2023**, *28*, 801. [CrossRef] [PubMed]
- [7] Diaz, A.H.; Duque-Noreña, M.; Chamorro, E. Unveiling an Electronic LogP Analogue Within the Conceptual Density Functional Theory Framework. *Chem. Phys.* **2024**, *584*, 112346. [CrossRef]
- [8] Thompson, J.; Cramer, C.; Truhlar, D. Predicting Aqueous Solubilities from Aqueous Free Energies of Solvation and Experimental or Calculated Vapor Pressures of Pure Substances. *J. Chem. Phys.* **2003**, *119*, 1661–1670. [CrossRef]
- [9] Connors, K.A.; Mecozzi, S. *Thermodynamics of Pharmaceutical Systems*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2010. [CrossRef]
- [10] Letcher, T.M. *Thermodynamics, Solubility and Environmental Issues*; Elsevier Science and Technology Books: Amsterdam, The Netherlands, 2007. [CrossRef]
- [11] Chung, Y.; Vermeire, F.H.; Wu, H.; Walker, P.J.; Abraham, M.H.; Green, W.H. Group Contribution and Machine Learning Approaches to Predict Abraham Solute Parameters, Solvation Free Energy, and Solvation Enthalpy. *J. Chem. Inf. Model.* **2022**, *62*, 433–446. [CrossRef] [PubMed]
- [12] Delgado, E.J.; Jaña, G.A. Quantitative Prediction of Solvation Free Energy in Octanol of Organic Compounds. *Int. J. Mol. Sci.* **2009**, *10*, 1031–1044. [CrossRef]
- [13] Yu, X.; Wang, H.; Acree, W., Jr.; Deng, J. QSPR Models for Solvation Enthalpy Based on Quantum Chemical Descriptors. *J. Mol. Liq.* **2023**, *389*, 122884. [CrossRef]
- [14] Sharma, P.; Ranjan, P.; Chakraborty, T. Applications of Conceptual Density Functional Theory in Reference to Quantitative Structure–Activity/Property Relationship. *Mol. Phys.* **2024**, *122*, e2331620. [CrossRef]
- [15] Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics (Volume I and II)*; WILEY-VCH Verlag GmbH & Co., KGa: Weinheim, Germany, 2009; Available online: <https://www.wiley.com/en-us/shop/general-chemistry/molecular-descriptors-for-chemoinformatics-volume-i-alphabetical-listing-volume-ii-appendices-references-2nd-revised-and-enlarged-edition-p-9783527628773> (accessed on 1 February 2026).
- [16] Sodaei, Z.; Ekrami, S.; Hashemianzadeh, S.M. Machine Learning Analysis of Molecular Dynamics Properties Influencing Drug Solubility. *Sci. Rep.* **2025**, *15*, 26955. [CrossRef] [PubMed]
- [17] Egert, T.; Langowski, H.-C. Linear Solvation Energy Relationships (LSERs) for Robust Prediction of Partition Coefficients Between Low Density Polyethylene and Water. Part II: Model Evaluation and Benchmarking. *Eur. J. Pharm. Sci.* **2022**, *172*, 106138. [CrossRef] [PubMed]
- [18] Sun, Y.; Hou, T.; He, X.; Man, V.H.; Wang, J. Development and Test of Highly Accurate Endpoint Free Energy Methods. 2: Prediction of Logarithm of n-Octanol–Water Partition Coefficient (logP) for Druglike Molecules Using MM-PBSA Method. *J. Comput. Chem.* **2023**, *44*, 1300–1311. [CrossRef] [PubMed]
- [19] Brown, T.N. QSPRs for Predicting Equilibrium Partitioning in Solvent–Air Systems from the Chemical Structures of Solutes and Solvents. *J. Solut. Chem.* **2022**, *51*, 1101–1132. [CrossRef]
- [20] Isert, C.; Kromann, J.; Stiefl, N.; Schneider, G.; Lewis, R. Machine Learning for Fast, Quantum Mechanics-Based Approximation of Drug Lipophilicity. *ACS Omega* **2023**, *8*, 2046–2056. [CrossRef] [PubMed]
- [21] Baldridge, K.; Klamt, A. First Principles Implementation of Solvent Effects Without Outlying Charge Error. *J. Chem. Phys.* **1997**, *106*, 6622–6633. [CrossRef]
- [22] Tshilande, N.; Mammino, L.; Bilonde, M.K. The Study of Molecules and Processes in Solution: An Overview of Questions, Approaches and Applications. *Computation* **2024**, *12*, 78. [CrossRef]
- [23] Liu, S.-C.; Zhu, X.-R.; Liu, D.-Y.; Fang, D.-C. DFT Calculations in Solution Systems: Solvation Energy, Dispersion Energy And Entropy. *Phys. Chem. Chem. Phys.* **2023**, *25*, 913–931. [CrossRef] [PubMed]
- [24] Tomasi, J.; Persico, M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994**, *94*, 2027–2094. [CrossRef]
- [25] Barone, V.; Cossi, M.; Tomasi, J. A New Definition of Cavities for the Computation of Solvation Free Energies by the Polarizable Continuum Model. *J. Chem. Phys.* **1997**, *107*, 3210–3221. [CrossRef]
- [26] Mennucci, B.; Tomasi, J. Continuum Solvation Models: A New Approach to the Problem of Solute’s Charge Distribution and Cavity Boundaries. *J. Chem. Phys.* **1997**, *106*, 5151–5158. [CrossRef]
- [27] Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001. [CrossRef]
- [28] Wittmann, L.; Selzer, C.; Grimme, S. A Diverse and Chemically Relevant Solvation Model Benchmark Set with Flexible Molecules and Conformer Ensembles. *Chem. Sci.* **2025**, *16*, 22976–22995. [CrossRef] [PubMed]

- [29] Thompson, J.D.; Cramer, C.J.; Truhlar, D.G. New Universal Solvation Model and Comparison of the Accuracy of the SM5.42R, SM5.43R, C-PCM, D-PCM, and IEF-PCM Continuum Solvation Models for Aqueous and Organic Solvation Free Energies and for Vapor Pressures. *J. Phys. Chem. A* **2004**, *108*, 6532–6542. [CrossRef]
- [30] Bell, I.H.; Mickoleit, E.; Hsieh, C.-M.; Lin, S.-T.; Vrabec, J.; Breitkopf, C.; Jäger, A. A Benchmark Open-Source Implementation of COSMO-SAC. *J. Chem. Theory Comput.* **2020**, *16*, 2635–2646. [CrossRef] [PubMed]
- [31] Marenich, A.V.; Olson, R.M.; Kelly, C.P.; Cramer, C.J.; Truhlar, D.G. Self-Consistent Reaction Field Model for Aqueous and Nonaqueous Solutions Based on Accurate Polarized Partial Charges. *J. Chem. Theory Comput.* **2007**, *3*, 2011–2033. [CrossRef] [PubMed]
- [32] Maldonado, A.M.; Hagiwara, S.; Choi, T.H.; Eckert, F.; Schwarz, K.; Sundaraman, R.; Otani, M.; Keith, J.A. Quantifying Uncertainties in Solvation Procedures for Modeling Aqueous Phase Reaction Mechanisms. *J. Phys. Chem. A* **2021**, *125*, 154–164. [CrossRef] [PubMed]
- [33] Uhlig, H. The Solubilities of Gases and Surface Tension. *J. Phys. Chem.* **1937**, *41*, 1215–1226. [CrossRef]
- [34] Pierotti, R.A. The Solubility of Gases in Liquids ¹. *J. Phys. Chem.* **1963**, *67*, 1840–1845. [CrossRef]
- [35] Pierotti, R.A. Aqueous Solutions of Nonpolar Gases ¹. *J. Phys. Chem.* **1965**, *69*, 281–288. [CrossRef]
- [36] Pierotti, R.A. A Scaled Particle Theory of Aqueous and Nonaqueous Solutions. *Chem. Rev.* **1976**, *76*, 717–726. [CrossRef]
- [37] Floris, F.; Tomasi, J. Evaluation of the Dispersion Contribution to the Solvation Energy. A Simple Computational Model in the Continuum Approximation. *J. Comput. Chem.* **1989**, *10*, 616–627. [CrossRef]
- [38] Amovilli, C. Calculation of the Dispersion Energy Contribution to the Solvation Free Energy. *Chem. Phys. Lett.* **1994**, *229*, 244–249. [CrossRef]
- [39] Amovilli, C.; Mennucci, B. Self-Consistent-Field Calculation of Pauli Repulsion and Dispersion Contributions to the Solvation Free Energy in the Polarizable Continuum Model. *J. Phys. Chem. B* **1997**, *101*, 1051–1057. [CrossRef]
- [40] Sepassi, K.; Yalkowsky, S. Solubility Prediction in Octanol: A Technical Note. *AAPS PharmSciTech.* **2006**, *7*, 26. [CrossRef] [PubMed]
- [41] Bode, B.M.; Gordon, M.S. Macmolplt: A Graphical User Interface for GAMESS. *J. Mol. Graph. Model.* **1998**, *16*, 133–138. [CrossRef] [PubMed]
- [42] Vosko, S.; Wilk, L.; Nusair, M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200. [CrossRef]
- [43] Perdew, J.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868. [CrossRef] [PubMed]
- [44] Šebesta, F.; Sovová, Z.; Burda, J.V. Determination of Amino Acids' pKa: Importance of Cavity Scaling within Implicit Solvation Models and Choice of DFT Functionals. *J. Phys. Chem. B* **2024**, *128*, 1627–1637. [CrossRef] [PubMed]
- [45] Ben-Naim, A.; Marcus, Y. Solvation Thermodynamics of Nonionic Solutes. *J. Chem. Phys.* **1984**, *81*, 2016–2027. [CrossRef]
- [46] Silla, E.; Tuñón, I.; Pascual-Ahuir, J.L. GEPOL: An Improved Description of Molecular Surfaces II. Computing the Molecular Area and Volume. *J. Comput. Chem.* **1991**, *12*, 1077–1088. [CrossRef]
- [47] Pascual-Ahuir, J.L.; Silla, E.; Tomasi, J.; Bonaccorsi, R. Electrostatic Interaction of a Solute with a Continuum. Improved Description of the Cavity and of the Surface Cavity Bound Charge Distribution. *J. Comput. Chem.* **1987**, *8*, 778–787. [CrossRef]
- [48] Lide, D.R. (Ed.) *CRC Handbook of Chemistry and Physics*, 81st ed.; CRC Press LLC: Boca Raton, FL, USA, 2000. Available online: https://archive.org/details/crcheandbookofche0000unse_u9i8.
- [49] Yaws, C. *Yaws' Handbook of Thermodynamic and Physical Properties of Chemical Compounds*; McGraw-Hill: New York, NY, USA, 1997; Available online: <https://app.knovel.com/knovel2/Toc.jsp?BookID=667> (accessed on 1 February 2026).
- [50] Barca, G.M.J.; Bertoni, C.; Carrington, L.; Datta, D.; De Silva, N.; Deustua, J.E.; Fedorov, D.G.; Gour, J.R.; Gunina, A.O.; Guidez, E.; et al. Recent Developments in the General Atomic and Molecular Electronic Structure System. *J. Chem. Phys.* **2020**, *152*, 154102. [CrossRef] [PubMed]
- [51] Bensberg, M.; Tütscher, P.; Unsleber, J.P.; Reiher, M.; Neugebauer, J. Solvation Free Energies in Subsystem Density Functional Theory. *J. Chem. Theory Comput.* **2022**, *18*, 723–740. [CrossRef] [PubMed]
- [52] Rizzo, R.C.; Aynechi, T.; Case, D.; Kuntz, I. Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions. *J. Chem. Theory Comput.* **2006**, *2*, 128–139. [CrossRef] [PubMed]
- [53] Marenich, A.V.; Cramer, C.J.; Truhlar, D.G. Performance of SM6, SM8, and SMD on the SAMPL1 Test Set for the Prediction of Small-Molecule Solvation Free Energies. *J. Phys. Chem. B* **2009**, *113*, 4538–4543. [CrossRef] [PubMed]
- [54] Wang, J.; Wang, W.; Huo, S.; Lee, M.; Kollman, P.A. Solvation Model Based on Weighted Solvent Accessible Surface Area. *J. Phys. Chem. B* **2001**, *105*, 5055–5067. [CrossRef]
- [55] Marenich, A.V.; Cramer, C.J.; Truhlar, D.G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic

- Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396. [[CrossRef](#)] [[PubMed](#)]
- [56] Raevsky, O.A.; Grigor'ev, V.J.; Raevskaja, O.E.; Schaper, K.-J. Physicochemical Properties/Descriptors Governing the Solubility and Partitioning of Chemicals in Water-Solvent-Gas Systems. Part 1. Partitioning Between Octanol and Air. *SAR QSAR Environ. Res.* **2006**, *17*, 285–297. [[CrossRef](#)] [[PubMed](#)]
- [57] Hou, T.J.; Xu, X.J. ADME Evaluation in Drug Discovery. 2. Prediction of Partition Coefficient by Atom-Additive Approach Based on Atom-Weighted Solvent Accessible Surface Areas. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1058–1067. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the publisher or editors. The publisher and editors assume no responsibility for any injury or damage resulting from the use of information contained herein.